**DATASET COMPARISON**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr. No.** | **Dataset (Source)** | **Shape (rows, columns)** | **Features (columns)** | **Accuracy achieved** | **ROC-AUC Score** |
| 1. | Cardiovascular Diseases Risk Prediction Dataset (2021 BRFSS Dataset from CDC available on Kaggle)  <https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset>  [Dataset Name: CVD\_cleaned.csv] | (308854, 19) | * General\_Health * Checkup * Exercise * **Heart\_Disease** *(target)* * Skin\_Cancer * Other\_Cancer * Depression * Diabetes * Arthritis * Sex * Age\_Category * Height\_(cm) * Weight\_(kg) * BMI * Smoking\_History * Alcohol\_Consumption * Fruit\_Consumption * Green\_Vegetables\_Consump ion * FriedPotato\_Consumption | 0.92 | 0.83 |
| 2. | Cardiovascular Disease Dataset – Mendeley Data (Doppala, Bhanu Prakash; Bhattacharyya, Debnath (2021), “Cardiovascular\_Disease\_Dataset”, Mendeley Data, V1, doi: 10.17632/dzz48mvjht.1)  <https://data.mendeley.com/datasets/dzz48mvjht/1>  [Dataset Name: Cardiovascular\_Disease\_Dataset.csv] | (1000, 14) | * patientid * age * gender * chestpain * restingBP * serumcholestrol * fastingbloodsugar * restingrelectro * maxheartrate * exerciseangia * oldpeak * slope * noofmajorvessels * **target** *(target)* | 0.97 | 1.00 |
| 3. | Heart Attack Dataset – Mendeley Data (Rashid, Tarik A.; Hassan, Bryar (2022), “Heart Attack Dataset”, Mendeley Data, V1, doi: 10.17632/wmhctcrt5v.1)  <https://data.mendeley.com/datasets/wmhctcrt5v/1>  Also available on Kaggle: <https://www.kaggle.com/datasets/subhajeetdas/mendeley-heart-attack-dataset>  [Dataset Name: Medicaldataset.csv] | (1319, 9) | * Age * Gender * Heart rate * Systolic blood pressure * Diastolic blood pressure * Blood sugar * CK-MB * Troponin * **Result** *(target)* | 0.98 | 0.99 |
| 4. | Cardiovascular Disease Dataset [3 datasets]  (Rajib Kumar Halder, November 10, 2020, "Cardiovascular Disease Dataset", IEEE Dataport, doi: <https://dx.doi.org/10.21227/7qm5-dz13>.)   1. Cardio Data (Kaggle Dataset)   [Dataset Name: Cardio\_Data.csv] | (70000,  12) | * age * gender * height * weight * ap\_hi * ap\_lo * cholesterol * gluc * smoke * alco * active * **target** *(target)* | 0.74 | 0.80 |
| 5. | Cardiovascular Disease Dataset [3 datasets]  (Rajib Kumar Halder, November 10, 2020, "Cardiovascular Disease Dataset", IEEE Dataport, doi: <https://dx.doi.org/10.21227/7qm5-dz13>.)   1. Hungarian (UCI Machine Learning Repository)   [Dataset Name: Hungarian\_Data.csv] | (294, 14) | * age * sex * cp * trestbps * chol * fbs * restecg * thalach * exang * oldpeak * slope * ca * thal * **target** *(target)* | 0.66 | 0.78 |
| 6. | Cardiovascular Disease Dataset [3 datasets]  (Rajib Kumar Halder, November 10, 2020, "Cardiovascular Disease Dataset", IEEE Dataport, doi: <https://dx.doi.org/10.21227/7qm5-dz13>.)   1. Cleveland (UCI Machine Learning Repository)   [Dataset Name: Cleveland\_Data.csv] | (303, 14) | * age * sex * cp * trestbps * chol * fbs * restecg * thalach * exang * oldpeak * slope * ca * thal * **target** *(target)* | 0.49 | 0.71 |